Dynamics and Statistics of the Climate System, 2016, 1–21
doi:10.1093/climsys/dzx001
Advance Access Publication Date: February 15, 2017
Research Article

Research Article

State dependence of climate sensitivity: attractor constraints and palaeoclimate regimes

Anna S von der Heydt[a,*] and Peter Ashwin[b]

[a]Institute for Marine and Atmospheric Research, Center for Extreme Matter and Emergent Phenomena, Utrecht University, Utrecht, The Netherlands and [b]Centre for Systems, Dynamics and Control, Department of Mathematics, University of Exeter, Exeter EX4 4QF, UK.

*Correspondence Anna S von der Heydt, Institute for Marine and Atmospheric Research, Center for Extreme Matter and Emergent Phenomena, Utrecht University, Princetonplein 5, 3584CC Utrecht, The Netherlands; E-mail: A.S.vonderHeydt@uu.nl

Received 31 March 2016; Revised 13 January 2017; Accepted 17 January 2017

Abstract

Equilibrium climate sensitivity is a key predictor of climate change. However, it is not very well constrained, either by climate models or by observational data. The reasons for this include strong internal variability and forcing on many timescales. In practice, this means that the 'equilibrium' will only be relative to fixing the *slow* feedback processes before comparing palaeoclimate sensitivity estimates with estimates from model simulations. In addition, information from the late Pleistocene ice age cycles indicates that the climate cycles between cold and warm regimes, and the climate sensitivity varies considerably between regime because of *fast* feedback processes changing relative strength and timescales over one cycle. In this paper, we consider climate sensitivity for quite general climate dynamics. Using a conceptual Earth system model of Gildor andTziperman (A sea ice climate switch mechanism for the 100-kyr glacial cycles. *J Geophys Res* 2001; 106: 9117–33) (with Milankovich forcing and dynamical ocean biogeochemistry), we explore various ways of quantifying the state dependence of climate sensitivity from unperturbed and perturbed model time series. Even without considering any perturbation, we suggest that climate sensitivity can be usefully thought of as a distribution that quantifies variability within the 'climate attractor'. On the 'climate attractor', there is a strong dependence on climate state or more specifically on the 'climate regime' where fast processes are approximately in equilibrium. We also consider perturbations by instantaneous doubling of CO_2 and similarly find a strong dependence on the climate state using our approach.

Key words: Climate response to perturbations, climate sensitivity, conceptual climate models, glacial–interglacial cycles, palaeoclimate.

1. Introduction

In order to estimate the anthropogenic impact on climate in the future, the response of the climate system to the present perturbation by greenhouse gases needs to be quantified. A frequently used measure for the response to changes in atmospheric CO_2 concentration is the equilibrium climate sensitivity (ECS). This is defined as the increase in the

© The Author 2017. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

global mean of the Earth's surface temperature per radiative forcing change after the fast-acting feedback processes in the Earth system have come into equilibrium (Charney, 1979). In the IPCC literature, ECS is frequently given as temperature increase per CO_2 doubling, i.e. in units of K, while the equilibrium climate sensitivity parameter describes the warming per radiative forcing, i.e., in units of K $(W\ m^{-2})^{-1}$. Here, we use the term ECS for both quantities and mostly use the units of warming per radiative forcing. Frequently, it is estimated by climate model simulations, where the atmospheric CO_2 concentration is doubled within a few decades, and equilibrium is assumed after typically 100–200 years; slow climate processes are kept stationary (and non-dynamic) in these model simulations. In (IPCC 2013), regression of the temperature change versus net radiative imbalance at the top of the atmosphere (Gregory *et al.*, 2004) is used to estimate ECS. ECS is the benchmark quantity for climate models but is still characterized by a considerable uncertainty of 1.5–4.5 K per CO_2 doubling (IPCC, 2013) and neither recent observations have narrowed down the range of expected climate change (Knutti and Hegerl, 2008). It is also clear that the feedbacks, and hence the ECS, will depend on climate state (e.g. Senior and Mitchell, 2000; Gregory *et al.*, 2004; Crucifix, 2006; Andrews and Forster, 2008; Yoshimori *et al.*, 2011; Caballero and Huber, 2013; von der Heydt *et al.*, 2014). Palaeoclimate studies have tried to use proxy records to independently constrain ECS (Rohling *et al.*, 2012), but even by taking account of fast feedback processes that depend on the climate state (von der Heydt *et al.*, 2014, Köhler *et al.*, 2015), it remains difficult to further constrain the range of expected climate warming. In particular, temperature changes considerably larger than the mean value of 3 K per CO_2 doubling (IPCC, 2013) as a consequence of atmospheric CO_2 increase cannot be excluded.

The observed warming of the Earth involves both direct radiative forcing and a variety of (positive and negative) fast feedback mechanisms. These are mostly related to atmospheric water vapour content, sea ice and cloud albedo and aerosol concentrations. On the slower (decadal) timescales, ocean heat uptake also contributes to the radiative (im-)balance. Quantifying the (fast) forcing is therefore not an easy task and limits our ability to reduce the uncertainty on climate sensitivity (Schwartz, 2012). Moreover, the internal variability of the climate system on many timescales adds another type of uncertainty to the value of ECS because assumptions in the definition of ECS such as the timescale separation may not be met. In fact, the most appropriate definition of ECS in the presence of natural variability and forcing is still under debate. Ghil *et al.* propose a non-autonomous stochastic approach in terms of random dynamical systems (Ghil, 2016; Chekroun *et al.*, 2011) and suggest that climate sensitivity corresponds to a derivative of a metric (Wasserstein distance) evaluated for the invariant measure with respect to some parameter that is changed. Other approaches include considering perturbations that are not necessarily in the linear regime: Dijkstra and Viebahn (2015) use a conditional non-linear optimization approach to define climate sensitivity.

In this paper, we discuss climate sensitivity as a property of the climate dynamics projected into the space of forcing R to global mean temperature T. In particular, section 2 discusses ECS for the unperturbed system with slow variability in terms of pairs of points on or near the 'climate attractor'. We relate this to more usual concepts of ECS and use this as a way to discuss state dependence. In particular, we discuss 'climate regimes' such that the sensitivity is well constrained within a regime, but poorly constrained while switching between regimes. We also discuss possible perturbations and differentiate those that give return to the same climate attractor from those that do not. In Section 3, we explore these ideas using a specific conceptual model of the Earth system (Gildor and Tziperman, 2001; Gildor *et al.*, 2002) that includes dynamic CO_2 and is able to simulate glacial–interglacial cycles as relaxation oscillations. In this section, the model is also perturbed in various ways to obtain (state-dependent) distributions of climate sensitivity. We conclude in section 4 with a discussion of some issues that arise related to the extraction and interpretation of ECS distributions from palaeoclimate records.

2. Climate sensitivity and dynamics in (*T*,*R*) space

The usual approach to ECS is to consider the radiative energy balance for the global mean surface temperature T

$$\frac{dT}{dt} \sim R_f + R_{slow} + R_{fast} - R_{OLW}, \tag{1}$$

where R_f we understand as the radiation due to (external) forcings, e.g. the radiation received from the sun R_{ins} and the forcing R_{CO_2} due to the greenhouse effect of atmospheric CO_2. The fluxes R_{slow} and R_{fast} are contributions to the radiative balance due to a set of feedback processes in the climate system that are classified as slow and fast, respectively,

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

usually relative to the typical timescale of the forcing. Finally, $R_{OLW} = -\varepsilon\sigma_B T^4$ is the outgoing long-wave radiation determined by the Stefan-Boltzmann law (with σ_B the Stefan-Boltzmann constant and ε the emissivity of the atmosphere). The equation (1) suggests a simple (T,R) relationship, but in reality, it is hard to unpick this relationship, not least because fast feedbacks may potentially give multiple attractors when the slow feedbacks are fixed.

Here, we simply take the approach that T and R are quantities that one can (in principle) observe from a complex system, and we investigate the relationship between them. Given two climate states with forcings $R_{f,1}$ and $R_{f,2} = R_{f,1} + \Delta R$ and temperatures T_1 and $T_2 = T_1 + \Delta T$, we consider climate sensitivity as the ratio of temperature change to radiative forcing change (including slow feedbacks):

$$S = \frac{T_2 - T_1}{R_{f,2} + R_{slow,2} - R_{f,1} - R_{slow,1}} = \frac{\Delta T}{\Delta R_f + \Delta R_{slow}}. \tag{2}$$

If there is a functional relationship of the form $T = T(R)$ for $R = R_f + R_{slow}$ then in the limit of small differences in R we expect

$$S \approx \frac{dT}{dR} \tag{3}$$

In the (hypothetical) case that all slow processes are known and quantified and that a timescale separation between slow and fast processes exists, S is the usual ECS. However, we take the approach that equation (2) can still be studied when no functional relation $T(R)$ exists, and this naturally leads to distributions of ECS.

If we consider only CO_2 as forcing and the land ice-albedo feedback as slow process (i.e. $R_f + R_{slow,1} = R_{[CO_2,LI]}$), then the specific climate sensitivity is the ratio

$$S_{[CO_2,LI]} = \frac{T_2 - T_1}{R_{[CO_2,LI],2} - R_{[CO_2,LI],1}} = \frac{\Delta T}{\Delta R_{[CO_2,LI]}}. \tag{4}$$

A priori it is not clear how equation (4) depends on the two climate states being compared. Moreover, on the one hand, ΔR should be small to make a linear approximation valid. On the other hand, taking climate states where ΔR is large is more likely to give values that are insensitive to measurement errors, in particular for palaeoclimate records. Indeed, $S_{[CO_2,LI]}$ will only give a single value if T is a (smooth) function of $R_{[CO_2,LI]}$, and we consider asymptotically small ΔR: in this case, the distribution approaches a δ function centred at dT/dR.

2.1. Climate sensitivity on the climate attractor

Let us suppose that there is an attractor for the climate system that is stationary (this includes the possibility of a climate that is turbulent and/or that responds in a chaotic way to stationary quasi-periodic astronomical forcing) and that the system is on a trajectory that explores this attractor as time progresses. Comparing the climate states at times t_{ref} and $t_{ref} + \delta$, one can define climate sensitivity over a time interval $[t_{ref}, \delta]$ as

$$S_{[CO_2,LI]}(t_{ref}, \delta) = \frac{T(t_{ref} + \delta) - T(t_{ref})}{R_{[CO_2]}(t_{ref} + \delta) + R_{[LI]}(t_{ref} + \delta) - R_{[CO_2]}(t_{ref}) + R_{[LI]}(t_{ref})}. \tag{5}$$

By considering a range of possible reference times t_{ref} and delays δ, there will be a distribution of sensitivities.

An alternative approach is to assume there that is a stationary measure (or distribution) μ of points in the $(T, R_{[CO_2,LI]})$-plane weighted according to how often they are visited over asymptotically long times, i.e. we assume that

$$\mu(A) := \lim_{t' \to \infty} \frac{1}{t'} \ell\left(\{0 < s < t' : (T(t+s), R_{[CO_2,LI]}(t+s)) \in A\}\right). \tag{6}$$

is independent of t and typical initial condition, where $\ell(B)$ is the length of the set $B \subset \mathbb{R}$ (see Fig. 1a). The measure μ can be thought of as a projection of a natural measure on the attractor onto the two observables $(T, R_{[CO_2,LI]})$: it gives a distribution of associations between T and $R_{[CO_2,LI]}$.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

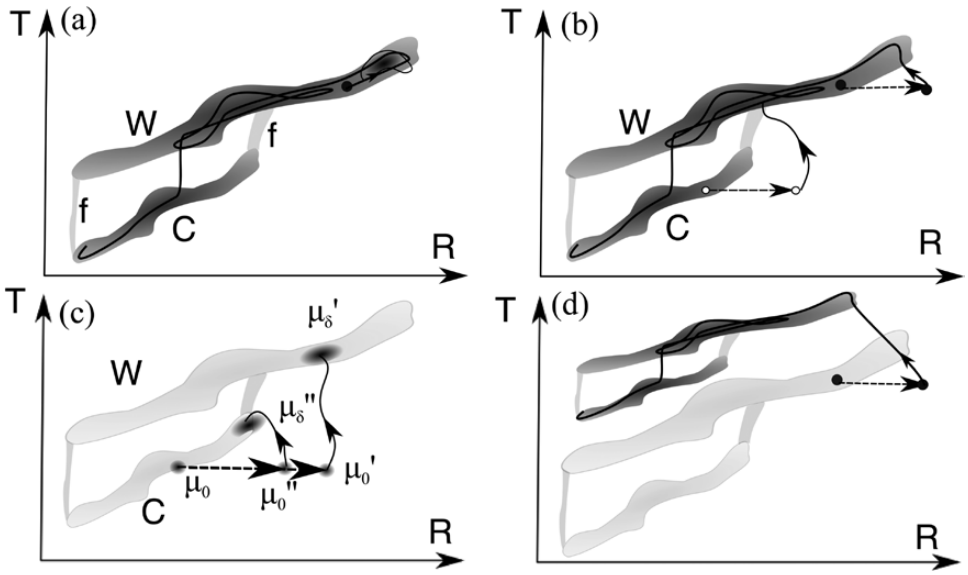

Figure 1. Schematic diagram showing global mean temperature T versus radiative forcing R due to atmospheric CO_2. (**a**) In the presence of natural forcing, we assume there is a stationary distribution μ (shown by grey scale) in the (T,R) plane—this is the projection of a dynamical measure onto this plane and can be divided into two climate regimes for the slow dynamics (shown here as C and W states), linked by fast changes (shown as f). Picking two points relative to this measure gives a distribution of slopes that quantifies long-term variability of climate sensitivity for this forcing. (**b**) A small impulsive change of R that does not structurally change the system (dashed line) takes the system state away from the attractor. If the perturbation does not change the attractor, after a transient (small arrow), we expect to continue to explore the plane according to the distribution μ. Depending on where the perturbation is applied and its size, the response may involve a switch between different regimes of the attractor (see the perturbation applied to C state). (**c**) A small impulsive change R (dashed line) moves an initial distribution μ_0 to a new location μ_0' or μ_0'' away from the attractor. After some time $\delta > 0$, we reach a perturbed distribution μ_δ' or μ_δ'': these may be in different regimes depending on the initial state and strength of the perturbation. (**d**) A large or structural change to the system will give a new attractor and a different set of asymptotic states.

The resulting distribution can be viewed as projection of an invariant measure on a climate attractor (Chekroun *et al.*, 2011) onto these observables and naturally leads to a distribution of climate sensitivities by picking pairs of points $(R_{[CO_2,LI]1,2}, T_{1,2})$ that are independently distributed according to μ and evaluating equation (4). In other words, for any (measurable) $A \subset \mathbb{R}$, we can use μ to assign a probability to the sensitivity being in A:

$$\text{Prob}(S_{[CO_2,LI]} \in A) := \mu \times \mu\left(\left\{(T_1, R_{[CO_2,LI],1}), (T_2, R_{[CO_2,LI],2}) : S_{[CO_2,LI]} \in A\right\}\right). \tag{7}$$

Note that equation (5) can be determined from a time series that does not necessarily explore the full attractor, while equation (7) considers states purely depending on the locations in the (T,R) plane. In Section 3, we give an example showing that the approaches (5) and (7) can give similar distributions when considering a wide range of δ and initial points.

2.2. Climate sensitivity, regimes and responses to perturbation

A study of palaeoclimate records, for example the ice age cycles of the last 800 kyr, shows the presence of markedly different 'regimes' of climate, namely periods of slowly varying climate and rapid transitions between these regimes (the deglaciations). We wish to evaluate the sensitivities associated within one regime and associated with changing regimes. For definiteness, we only consider climates with two regimes—a cold (C) and a warm (W) regime—in this paper. If one partitions the attractor into two regimes in state space, this implies a partition of μ into two distributions

$$\mu = \mu_C + \mu_W.$$

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

Evaluating the distribution of climate sensitivities corresponding to choosing typical endpoints relative to these distributions allows one to examine the sensitivities within regimes. In particular, one can define conditional distributions of sensitivities of the 'warm' (and similarly the 'cold') states by

$$\mathrm{Prob}(S^{WW}_{[CO_2,LI]} \in A) := \mu_W \times \mu_W\left(\left\{(T_1, R_{[CO_2,LI],1}), (T_2, R_{[CO_2,LI],2}) : S_{[CO_2,LI]} \in A\right\}\right) \tag{8}$$

where $S_{[CO_2,LI]}$ is as in equation (4). There are conditional sensitivities associated with regime changes, for example from C to W, this is

$$\mathrm{Prob}(S^{CW}_{[CO_2,LI]} \in A) := \mu_C \times \mu_W\left(\left\{(T_1, R_{[CO_2,LI],1}, T_2, R_{[CO_2,LI],2}) : S_{[CO_2,LI]} \in A\right\}\right) \tag{9}$$

and the distribution (7) can be thought of as the sum of the conditional distributions for S^{WW}, S^{CC}, S^{CW} and S^{WC}. Note that from the definition above

$$\mathrm{Prob}(S^{CW}_{[CO_2,LI]} \in A) = \mathrm{Prob}(S^{WC}_{[CO_2,LI]} \in A),$$

even if physically and when time progresses, the CW transition is different than the WC transition. For an optimal choice of regimes, one would aim to ensure that the distribution of sensitivities within each regime is tightly localized, while those associated with regime changes may be poorly localized. A regime could, therefore, be defined as a region in (T,R) space where the $T(R)$ relation is almost linear.

We now consider response to two types of instantaneous perturbation. The first type of perturbation does not structurally change the system or leave the basin of the current attractor: the response shows transient decay back to the attractor followed by continued motion on the same attractor. The response to such a perturbation includes the possibility of switching between regimes of the attractor, depending on the initial point on the attractor where the perturbation is applied. Figure 1b illustrates this schematically. In such a case, the distribution of sensitivities will potentially depend on the timescale δ of interest and the initial time t_{ref}. However, we expect it to decay to the (regime-dependent) sensitivity for large δ.

The second type of perturbation either structurally changes the system attractor or is large enough to place the state in the basin of a different attractor. In either case, the response will approach a new attractor as illustrated in Figure 1d. In this case, the distribution of sensitivities obtained by comparing initial and final states may not resemble regimes of either attractor.

Finally, we mention another approach to perturbation that is particularly useful for short-term prediction: Figure 1c starts with a localized (say Gaussian) distribution μ_0 centred on a perturbed reference state at some time t_{ref} and propagates this forwards to time $t_{ref} + \delta$ for some $\delta > 0$. The initial distribution will spread to give a localized measure μ_δ that gives a distribution of possible sensitivities via equation (4), which again may depend on the timescale δ, which again may depend on the timescale δ. ECS derived from palaeoclimate records typically reflects the climate sensitivity on the attractor (Fig. 1a), while model-determined ECS usually involves some type of perturbation (away from the attractor). In this sense, palaeo ECS and model ECS are conceptually different. In the next section, we explore how and when these two concepts can still give similar disributions, using a conceptual Earth system model.

3. Climate sensitivity in a conceptual climate model

The conceptual model of the climate system of (Gildor and Tziperman, 2001; Gildor *et al.*, 2002) has been shown to simulate the glacial–interglacial transitions; the model equations are given in Appendix A. In this model, the atmosphere is represented by four meridional boxes, while the ocean component consists of two layers of four meridional boxes each. The model includes land ice, sea ice and carbon-cycle effects, such that the atmospheric CO_2 concentration is a dynamic variable in the model. The model contains one dynamic fast feedback, namely the sea ice-albedo feedback evolving on (sub-)decadal timescales, and one slow feedback, the land ice-albedo feedback, which evolves on the order of millennial timescales. On the decade-to-century timescale the model includes an additional process in the surface radiative balance due to heat exchange between ocean and atmosphere. All other fast feedbacks (water vapour, clouds, aerosols, lapse rate) are represented by a fixed

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

temperature response to the radiative forcing in the system. In this case, as discussed in Appendix B, there is only one active slow feedback process, and the specific climate sensitivity parameter $S_{[CO_2,LI]}$ in equation (4) represents the model's ECS. Orbital forcing is included in the model through varying incoming solar radiation averaged over each atmospheric box on seasonal and orbital timescales and modulating the Northern Hemisphere land ice ablation term by the (northern polar box averaged) summer insolation on orbital timescales (Gildor and Tziperman, 2000).

The atmospheric CO_2 concentration in the full model system deserves further discussion because it can be viewed as both a forcing and a feedback. While the dynamic CO_2 is not essential for generating the glacial–interglacial cycles in the model, it feeds back on their amplitude; during cold periods when land ice is growing, reduced vertical mixing in the Southern Ocean and extended Southern Ocean sea ice cover leads to reduced atmospheric CO_2 (Gildor *et al.*, 2002). The exchange of CO_2 between ocean and atmosphere is fast (on timescales of a decade); however, the vertical mixing of surface-to-deep water masses in the Southern Ocean is affected by the temperature of the North Atlantic deep water, which evolves on slower timescales. The associated feedback process therefore acts on various timescales. When determining climate sensitivity, CO_2 is generally assumed a forcing. Here, it is important to keep in mind that it also can be viewed as a feedback as has been observed also in other models (Scheffer *et al.*, 2006) and will be the case in the real climate system, particularly on long (geological) timescales.

3.1. Glacial–interglacial cycles as relaxation oscillations

We first analyse a simulation with the climate model including prognostic pCO_2 and Milankovitch forcing. The simulation is started 500 kyr ago from initial conditions that are assumed to be close to the attractor and it is run up until the present day. The simulated glacial–interglacial cycles show a peak-to-peak global mean temperature difference of up to 4 K (Fig. 2a). Corresponding CO_2 differences are 75 ppmv, which here are completely generated by the effect of the solubility pump in the ocean.

In this model, the fast sea ice-albedo feedback is responsible for the abrupt glacial–interglacial variations—the so-called sea ice switch mechanism as suggested by Gildor and Tziperman (2001). The sea ice switch mechanism generates the glacial cycles in the model as self-sustained relaxation oscillations because the ice volume threshold for switching sea ice cover from 'on' to 'off' differs from the one for switching from 'off' to 'on' (Crucifix, 2012) (see Fig. 3a), and we use this to define two climate regimes, a cold *C* regime with extensive sea ice cover and a warm *W* regime without sea ice. When the land ice volume slowly grows (accumulation exceeds ablation), the atmospheric and surface ocean temperature decrease due to increasing albedo of the planet.

Once the polar surface ocean temperature has reached a critical value cold enough to form sea ice, the polar box is rapidly covered with sea ice, which further reduces the atmospheric temperature through the ice-albedo feedback and prevents evaporation from the polar ocean box (Fig. 2b). In addition, atmospheric moisture content is reduced due to lower temperatures, which leads to decreasing land ice volume (accumulation is smaller than ablation, Fig. 2c). Temperature starts rising again both due to smaller albedo and because the ocean warms below the insulating sea ice cover until it is warm enough to melt the polar sea ice, Fig. 2d). At this point, there is a change in regime: the global temperature quickly rises, moisture content in the atmosphere increases and the land ice starts growing again (accumulation becomes larger than ablation).

In this model, Milankovitch forcing is not necessary to generate the glacial–interglacial cycles, but it modifies them and makes them more irregular. Although there is some degree of synchronization of the glaciation and deglaciations to the orbital forcing, the relation between land ice and global mean solar radiation is not trivial (Fig. 2b). Milankovitch forcing mainly modulates the (otherwise constant) ablation of the Northern Hemisphere land ice, and therefore, while the land ice is growing and ocean temperatures decreasing in some (slightly warmer) periods, land ice accumulation becomes smaller than ablation and the ice growth and ocean cooling trends are episodically reversed (Fig. 2c).

3.2. Climate sensitivity for unperturbed climates

If one tries to determine climate sensitivity from past climate records, there is only one temporal realization of a trajectory on the climate attractor that can be measured: perturbations away from the attractor are not available. Defining the climate sensitivity in terms of the measure on the climate attractor (see Fig. 1a), we need to consider the relation between temperature T and radiative forcing due to CO_2 and land ice (the only slow process in the climate model). Figure 4a,b shows the probability density of (T,R) combinations for the 500 kyr trajectory discussed above, obtained by box-counting the frequency of visits to a uniform discretization of this range of T and R into 120×120

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

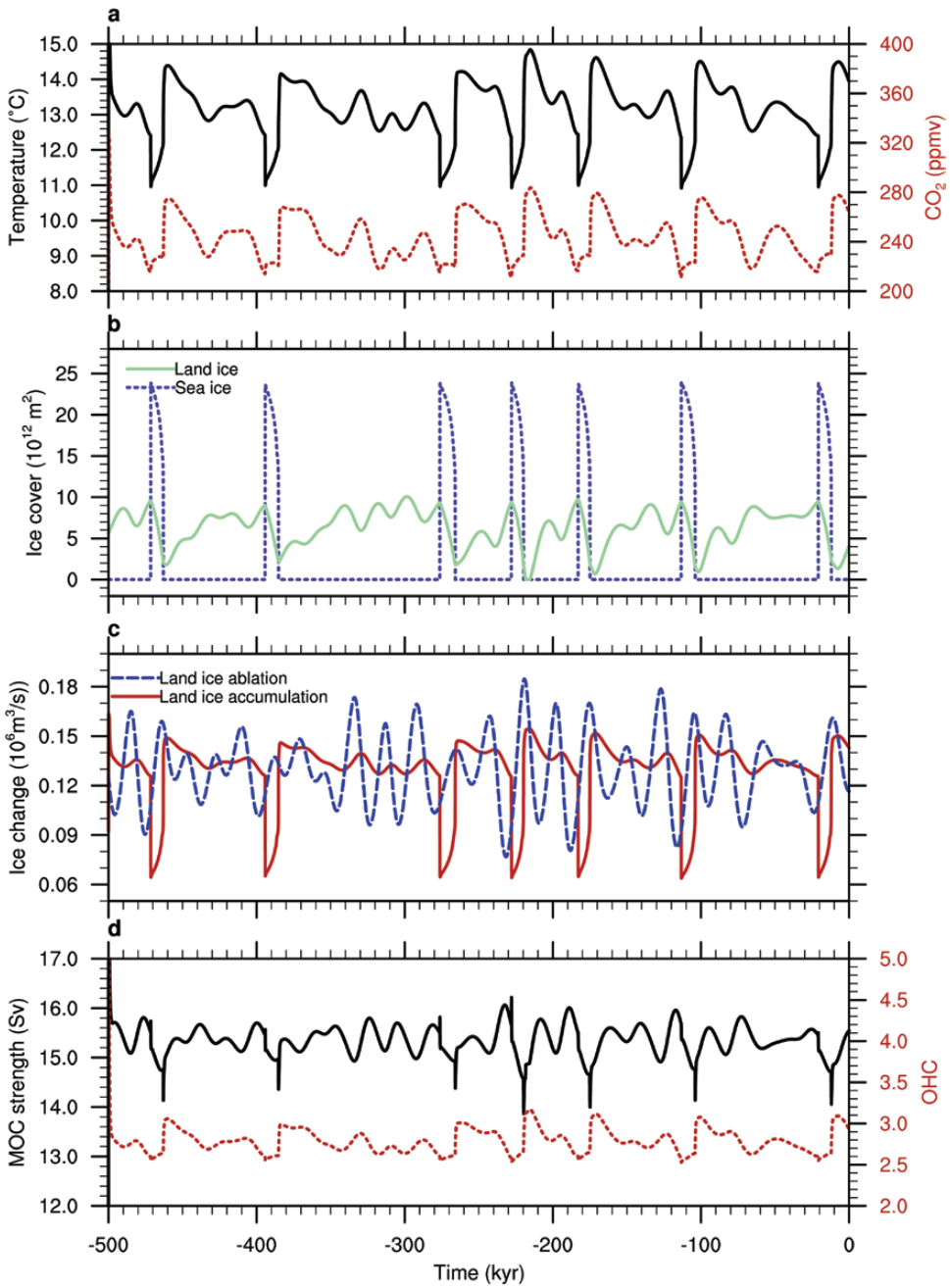

Figure 2. Glacial cycles of the box model, shown are time series of 100-year averages. (a) Simulated global mean surface temperature T (black line) and atmospheric CO_2 (red line); (b) land ice (green line) and sea ice (blue line) cover of the northern polar box; (c) land ice accumulation (red line) and ablation (blue line) in the northern polar box; (d) strength of the ocean meridional overturning circulation (black line), measured as the volume exchange between surface and deep northern polar ocean boxes and ocean heat content (red line).

cells (we remove a transient of length 10 000 years). This empirical distribution can be seen as an approximation of μ in equation (6).

In the special case of this relation being a linear function $T(R)$, the climate sensitivity is given by the slope of this line and constant for all climate states. However, it has been previously shown that in this climate model the climate sensitivity is strongly state dependent due to the fast sea ice- albedo feedback changing in strength between different climate

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

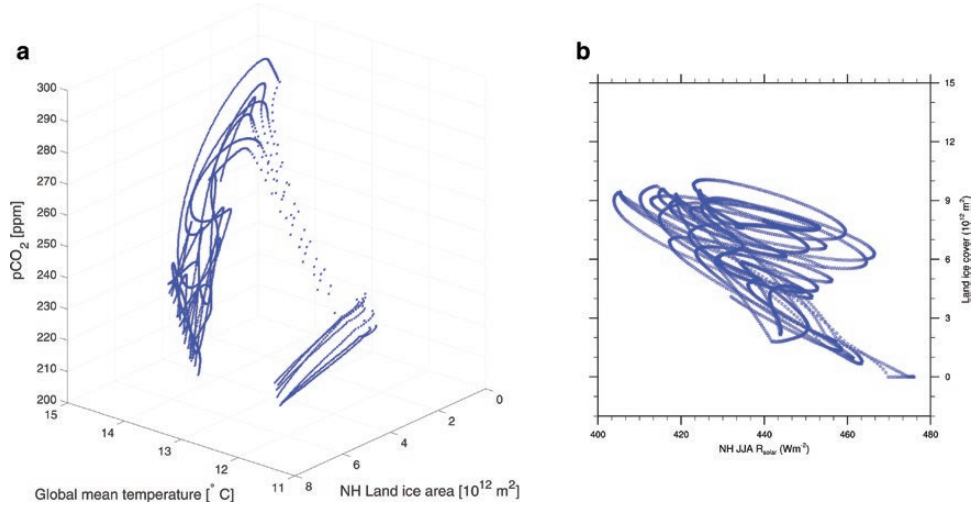

Figure 3. Phase diagrams of the glacial cycles of the box model; each point is a 100-year average. (a) Northern Hemisphere (NH) land ice cover as a function of global mean temperature and atmospheric CO_2; (b) NH land ice cover as a function of orbital variations in solar radiation R_{solar} defined as June–July–August (JJA) averaged insolation over the northern polar box of the model (45–90° averaged).

states (von der Heydt *et al.*, 2014), which allows the definition of a local climate sensitivity (for a reference climate state). Indeed, in Figure 4b, there appear to be regimes where the (T,R) relation is close to a (linear) function, but particularly, in the transition region from glacial to interglacial states, the sensitivity is less well defined or even negative.

From this data, we first estimate the slope of the relation S_{local} between T and $R_{[CO_2,LI]} = R_{[CO_2]} + R_{[LI]}$ (Fig. 4b) by a linear regression on the warm (W) and cold (C) parts of the data, giving $S^{W}_{local} = 0.45\text{K}\ (\text{W m}^{-2})^{-1}$ and $S^{C}_{local} = 0.54\ \text{K}\ (\text{W m}^{-2})^{-1}$, respectively. The linear regression on the warm (cold) part of the data gives an estimate of the mean of the distribution for the conditional sensitivity $S^{WW}_{[CO_2,LI]}$ ($S^{CC}_{[CO_2,LI]}$) as defined in equation (8). In the regression, all points with $T \leqslant 12°\text{C}$ are considered for S^{C}_{local} and points with $12.5°\text{C} \leqslant T \leqslant 14.5°\text{C}$ for S^{W}_{local}, respectively. Note that the temperature classification divides the data into climate states without Northern Hemisphere sea ice (W) and those where sea ice is present (C). The physical explanation for the higher sensitivity during the colder part of the data is that the presence of sea ice in the cold climate states leads to a stronger sea ice-albedo feedback.

As can be seen by the density of points in Figure 4b, even the almost linear parts of the (T,R) relation are not a (smooth) function: there is a distribution of slopes for each climate state. In Figure 5, $S_{[CO_2,LI]}(\delta, t_{ref})$ is shown for all values of t_{ref} and delays $\delta = 0 - \pm 25$ kyr (1/4 of the average period of the glacial–interglacial cycles), where the white (black) shading indicates very large ($\geqslant 3\ \text{K}\ (\text{W m}^{-2})^{-1}$) positive (all negative) values. A distribution of $S_{[CO_2,LI]}(\delta, t_{ref})$ is given in Figure 6a. We also classify $S_{[CO_2,LI]}(\delta, t_{ref})$ in terms of which regimes (C or W) are being compared at times T_{ref} and $T_{ref} + \delta$. The resulting distributions $S^{WW}_{[CO_2,LI]}(\delta, t_{ref})$, $S^{CC}_{[CO_2,LI]}(\delta, t_{ref})$ and $S^{CW/WC}_{[CO_2,LI]}(\delta, t_{ref})$ are shown in Figure 6b–d. Comparing only W states (without sea ice) leads to generally lower sensitivity, with its mean close to the value determined by the linear regression of only W states and a rather narrow distribution. Similarly, comparing only C states (with variable sea ice) results in somewhat higher sensitivities and a larger spread around the mean (which is again close to the linear regression of the C states). The larger spread is not surprising given that the (T,R) relation in the C regime is clearly non-linear.

The plots in Figure 6e–h correspond to Figure 6a–d but are calculated using equation (7) and the approximate measure μ whose density is shown in Figure 4b. Note the similarity of the distributions found by both methods. The largest (and most negative) values of $S_{[CO_2,LI]}$ in (Fig. 6a,e) originate from the cross-comparison of C and W regimes (Fig. 6d,h). On the other hand, the means of $S^{CC}_{[CO_2,LI]}$ agrees well with $S^{C}_{local} = 0.54\text{K}\ (\text{W m}^{-2})^{-1}$, while the mean of $S^{WW}_{[CO_2,LI]}$ agrees well with $S^{W}_{local} = 0.45\text{K}\ (\text{W m}^{-2})^{-1}$.

This suggests (i) the discretization used to approximate μ is sufficient to capture the main features of the regime-dependent sensitivity and (ii) that the reference times and delays considered sample the distribution μ well and so distributions of sensitivities given by equations (7) and (8) are similar.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

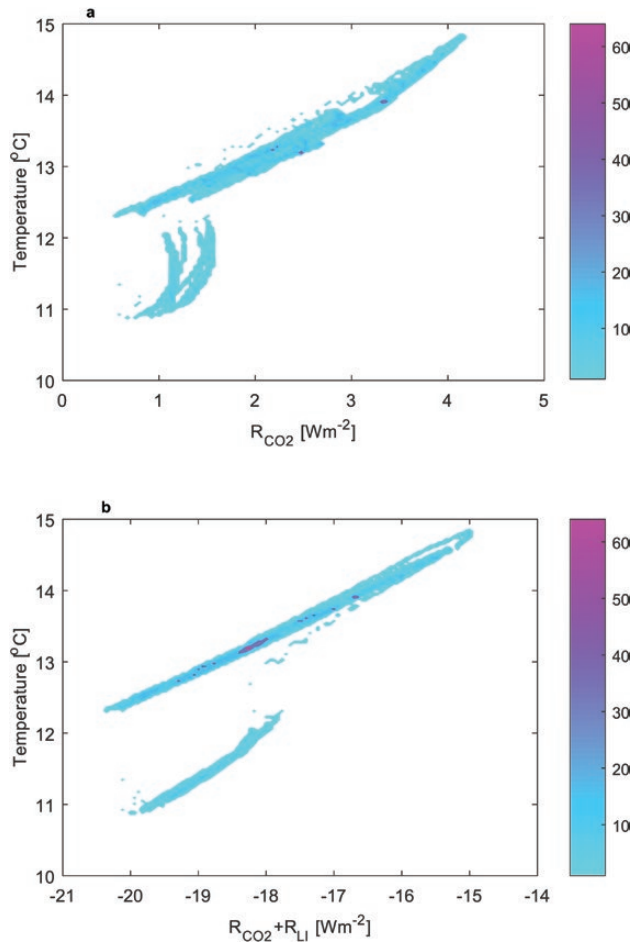

Figure 4. Relation between global mean temperature T and radiative contributions R due to pCO_2 $R_{[CO_2]}$ and land ice $R_{[LI]}$ (slow feedback). Colours (a.u.) show a box-counting approximation of the probability density distribution of the relation, after a transient has been removed (see text for details): (**a**) distribution of T versus R_{CO_2}; (**b**) distribution of T versus $R_{CO_2,LI}$. Observe the presence of two clear regions of high density: an upper regime of W = warm states and a lower regime of C = cold states where there is extensive sea ice present. Note the fast switches f between regimes contain very little density as they are much faster than the evolution within the W and C states.

3.3. Climate sensitivity for perturbed model climates

If we consider climate sensitivity as a local property of a natural measure on the climate attractor as illustrated in Figure 1a, we need to explore the set of points in the (T,R) plane that the model visits over the glacial–interglacial cycles. In climate models used for future prediction, the usual approach is to perturb the system in some way (e.g. double pCO_2) and study the response to this perturbation after some time. An initial distribution μ_0 evolves over a certain timescale δ after a perturbation away from the attractor has been applied as illustrated in Figure 1b,c. The distribution of sensitivities can be found by perturbing the system instantaneously, assuming that the perturbation is not too large and the system returns to the same attractor after a transient. Note, however, that this approach requires a different type of perturbation than what is usually applied in climate models; the standard procedure in general circulation models is to consider a prescribed (non-dynamic) CO_2 doubling as a perturbation, where in fact a different attractor than that of the full Earth system (including dynamic carbon cycle) is explored. In this section, we derive the model's ECS in response to a perturbation to the initial atmospheric CO_2 including a dynamic carbon cycle and evaluating the temperature response to this initial perturbation at different times (following the approach illustrated in Fig. 1c).

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

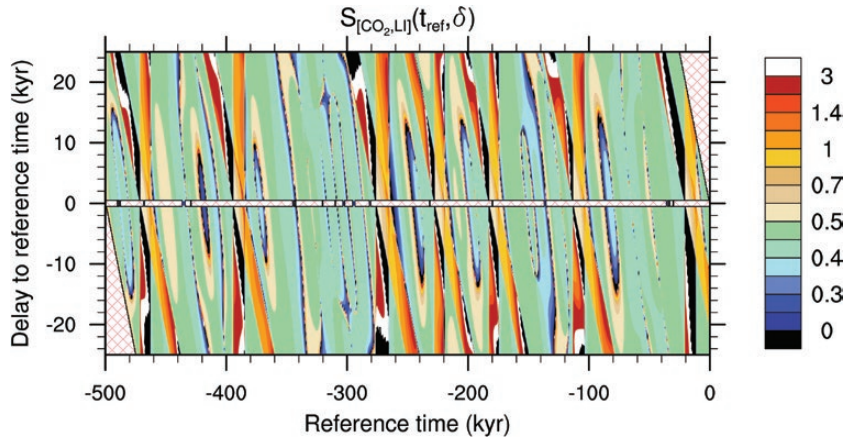

Figure 5. Equilibrium climate sensitivity approximated by the specific climate sensitivity $S_{[CO_2,LI]}(\delta, t_{ref})$ from equation (5) from a long glacial–interglacial simulation. All reference times along the time series are considered and delays to the reference time between $\delta = -25 \ldots + 25$ kyr (about half a glacial–interglacial period). The plot shows contours of $S_{[CO_2,LI]}(\delta, t_{ref})$ as a function of δ and t_{ref}. White shading indicates values of $S_{[CO_2,LI]}(\delta, t_{ref})$ 3 K (W m^{-2})$^{-1}$; black shading indicates negative values.

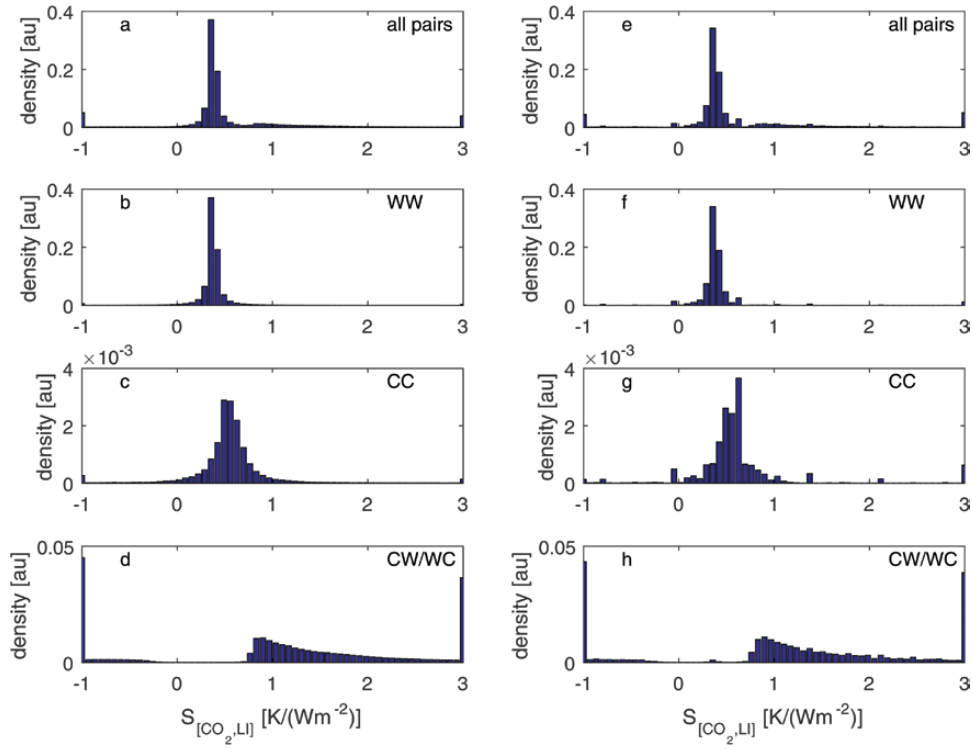

Figure 6. Distributions of the specific climate sensitivity $S_{[CO_2,LI]}$. The left panels (**a**– **d**) use equation (5) and a long glacial–interglacial simulation for a range of reference times and a range of delays greater than 500 yr. The right panels (**e**– **h**) use equation (4) and the approximation of μ in Figure 4b. Values of S outside the range [–1,3] are truncated to the endpoints of the domain. (a, e) All climate states are considered; (b, f) only pairs of climate states that are both in the *W* regime (no sea ice) are considered; (c, g) only pairs of climate states that are both in the *C* regime (sea ice present) are considered; (d, h) only pairs of climate states, where one is in the *W* regime and the other is in the *C* regime are considered. Observe that within each regime the distribution appears to be fairly tightly defined, while the *CW*/*WC* transitions have very long tails. Moreover, the distributions from the two methods give comparable results both within and across regimes.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

From the 500 kyr model time series shown in Figure 2, we chose two initial conditions, one in the *W* regime and one in the *C* regime with extensive sea ice. The CO_2 is doubled initially in the atmosphere and the extra CO_2 added by the doubling is uniformly subtracted from the ocean boxes, in order to conserve the total amount of carbon in the model system. Note that the atmospheric CO_2 in this model is purely determined by the biological pump in the oceans, while solubility effects play a minor role (Gildor *et al.*, 2002). Both perturbed initial conditions are run till present day (time 0), time series are shown in Figure 7 together with the unperturbed time series (as in Fig. 2). The time series starting from the *W* state quickly returns to the same temperature, CO_2, sea ice and land ice time series (dotted lines in Fig. 7), where glacial inceptions and deglaciations occur at exactly the same time as in the unperturbed simulation (thin solid lines). In contrast, the perturbed time series starting from the *C* state (dashed lines) does not return to the same unperturbed time series; the sea ice present initially is melted by the initial warming, and the system undergoes a transition to the *W* regime. While on the long term, the variation in temperature, CO_2, sea ice and land ice covers the same range as in the unperturbed simulation, the timing of glacial inceptions and deglaciations is different from the unperturbed simulation. This suggests that the applied perturbation is indeed small enough such that the system returns to the same attractor as can be seen in the left panel of Figure 8. When the initial condition lies in the *C* regime, the perturbation induces a regime switch, after which the same attractor is explored, but on a different trajectory.

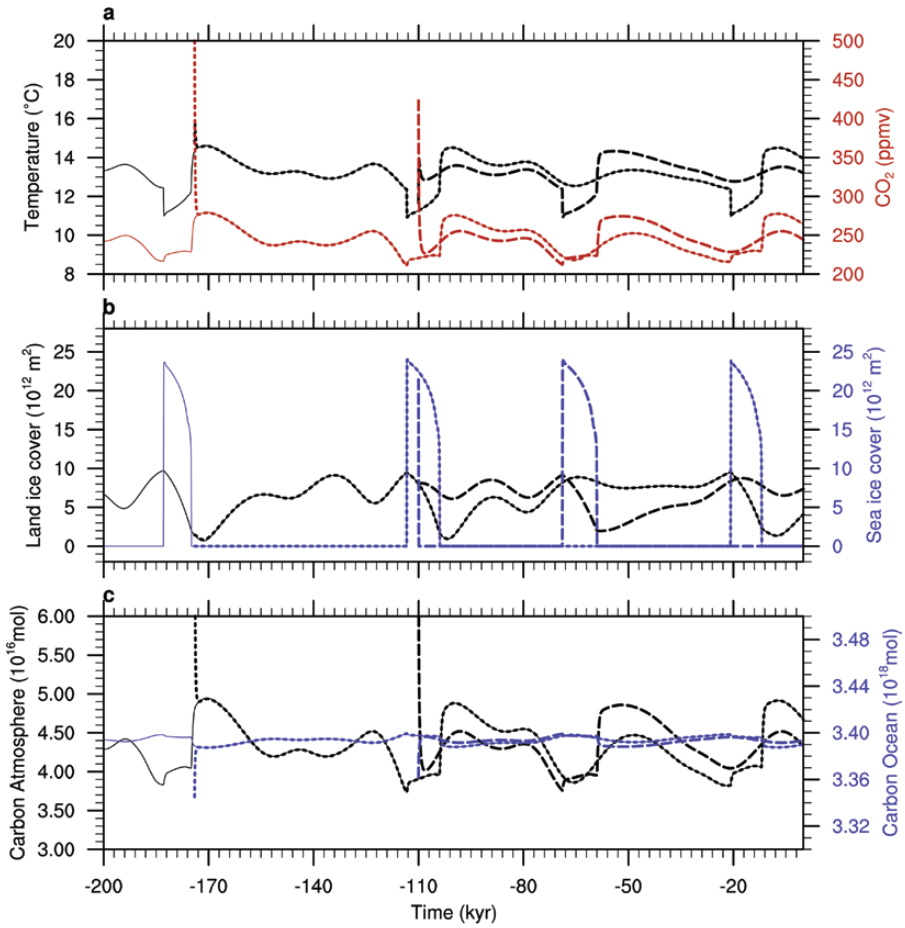

Figure 7. Perturbed glacial cycles of the box model, shown are time series of 100-year averages. The perturbation consists of doubling the atmospheric CO_2 initially, while compensating for the added carbon in the ocean model. Perturbations are applied at two instances: during an interglacial (dotted lines) and during a glacial with extensive sea ice cover (dashed lines). (**a**) Simulated global mean surface temperature *T* (black lines) and atmospheric CO_2 (red lines); (**b**) land ice (black lines) and sea ice (blue lines) cover of the northern polar box; (**c**) total carbon in the atmosphere (black lines) and the ocean (blue lines).

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

We have also applied a modified perturbation, where the total amount of carbon is not conserved; atmospheric CO_2 is doubled, while the oceanic CO_2 is unchanged. In this case, the result of the perturbation is to shift the attractor towards higher temperatures, higher atmospheric CO_2 and slightly different amounts of land ice, as shown in the right panel of Figure 8. This type of perturbation might reflect more realistically the present-day climate change situation assuming that the carbon injected into the system originates from a geological reservoir. However, while the attractor remains very similar in shape in this case but is shifted in phase space, other components of the model system such as the land ice might need adaptations of their parameters. The situation reflects the one depicted in Figure 1d and we will not further discuss the response to this type of perturbation but instead focus on the situation, where the climate system returns to the same attractor after the perturbation (Fig. 8, left panel).

Starting from 250 different initial conditions chosen along the 500 kyr time series shown in Figure 2a (one initial condition every 2000 years), the model is integrated for 500 years to give control runs of temperature $T^{cntrl(i)}(t)$ and radiative forcing time series $R^{cntrl(i)}_{[CO_2]}(t) + R^{cntrl(i)}_{[LI]}(t)$, respectively, where the index $i = 1,...,250$ denotes the initial condition. The initial CO_2 concentration pCO_2^0 in these simulations varies between 210 and 290 ppm, while the global mean temperature varies between 10.8 and 14.9°C. A second set of simulations is performed, where the initial value of the CO_2 is doubled and then the model is integrated for 500 years, giving $T^{pert(i)}(t)$ and $R^{pert(i)}_{[CO_2]}(t) + R^{pert(i)}_{[LI]}(t)$. A (time-dependent) climate sensitivity is then determined from

$$S^{(i)}_{perturb}(t) = \frac{\Delta T^{(i)}(t)}{\Delta R^{(i)}(t)} = \frac{T^{pert(i)}(t) - T^{cntrl(i)}(t)}{R^{pert(i)}_{[CO_2]}(t) + R^{pert(i)}_{[LI]}(t) - R^{cntrl(i)}_{[CO_2]}(t) - R^{cntrl(i)}_{[LI]}(t)} \tag{10}$$

Figure 9 shows time series of CO_2, temperature, Northern Hemisphere sea ice cover, the ocean meridional overturning circulation strength and $S_{perturb}$ for a few of the ensemble members (both control and perturbed experiments). Clearly, the different timescales in the system become evident; the global mean atmospheric temperature reacts quickly to the elevated CO_2 level, and for those initial states that have sea ice, the sea ice melts within 10–20 years. As the CO_2 is dynamic in these simulations, the increased CO_2 gradient between ocean and atmosphere leads to a rather fast initial reduction in atmospheric CO_2 (timescale of ~ 10 years) (Gildor *et al.*, 2002), which then keeps decreasing on a longer timescale. After 500 years, temperature and CO_2 are almost back to their original values if the initial condition was within the *W* regime. However, the initial conditions within the *C* regime involve a regime shift and do not return to the same temperature and CO_2 level within 500 years. The strength of the meridional overturning circulation in the ocean weakly responds to the CO_2 perturbation on a slower timescale as can be seen in Figure 9d. The time-dependent

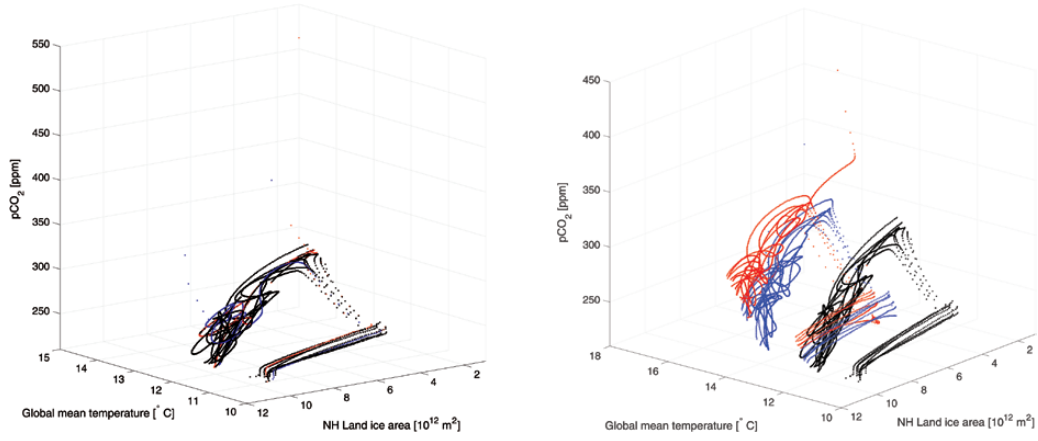

Figure 8. Climate attractor of the climate model showing responses to doubling CO_2 from states in the *C* regime (blue symbols) with and the *W* regime (red symbols). Black symbols represent the unperturbed attractor as shown in Figure 3a. We apply two different types of perturbations: (left panel) the total amount of carbon in the model system (ocean and atmosphere) is conserved. When doubling the CO_2 concentration in the atmosphere, the same amount of CO_2 is removed from the ocean. After an inital transient, the model returns to the same attractor as schematically illustrated in Figure 1b; (right panel) the atmospheric CO_2 is doubled without compensating in the ocean, meaning that extra carbon is added to the model system. In this case, the perturbed simulations return to an attractor that has a similar shape, but is shifted to higher CO_2 levels and global mean temperatures (see also Fig. 1d).

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

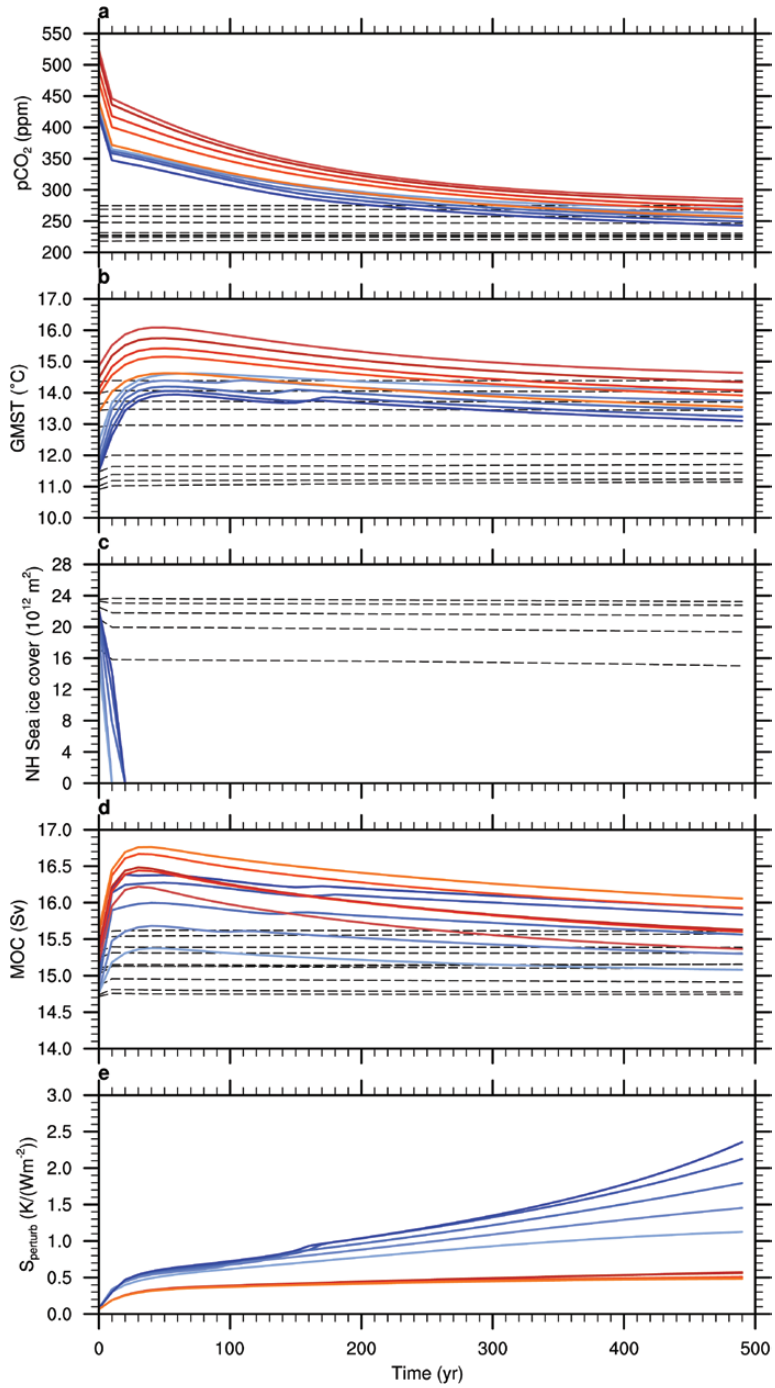

Figure 9. Climate sensitivity $S_{perturb}$ from perturbation experiments with dynamical CO_2. Shown are time series of experiments, where CO_2 is doubled initially and free to evolve until year 500 (coloured lines) along with the control experiments, where CO_2 is not doubled initially (black dashed lines). The ensemble starts from 250 initial conditions taken from the glacial–interglacial time series (500 kyr, shown in Fig. 2a). In this figure, we show 10 ensemble members, the blue lines have sea ice initially, while the red lines have no sea ice and darker red indicates warmer initial temperature. Climate sensitivity is determined following equation (10). (**a**) Atmospheric CO_2; (**b**) global mean surface temperature; (**c**) Northern Hemisphere sea ice fraction; (**d**) strength of the ocean meridional overturning circulation; (**e**) $S_{perturb}$.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

climate sensitivity $S_{perturb}$ is shown in Figure 9e; here, the different behaviour of the *C* and *W* states becomes particularly evident: while the response to the perturbation of the *W* states (red lines) seem to approach an 'equilibrium' value with some spread, increasing in time, the *C* states (blue lines) produce a wide range of responses depending on where the attractor is met after the perturbation.

Snapshots of distributions of $S_{perturb}$ are shown in Figure 10 for 100, 200 and 500 years after the perturbation. A fast-process equilibrium should be expected after 100–200 years; however, the spread in $S_{perturb}$ also increases with time, in particular for the *C* regime. $S_{perturb}$ of the *W* regime is similar to $S^{WW}_{[CO_2,LI]}$ after 100 and 200 years, but further spread out after 500 years. On the other hand, $S_{perturb}$ of the *C* states resembles $S^{CW}_{[CO2,LI]}$ already after 200 years, and afterwards spreads out even further. In this ensemble, $S^{CC}_{[CO2,LI]}$ is in fact never observed because the perturbation always induces a *C*–*W* transition.

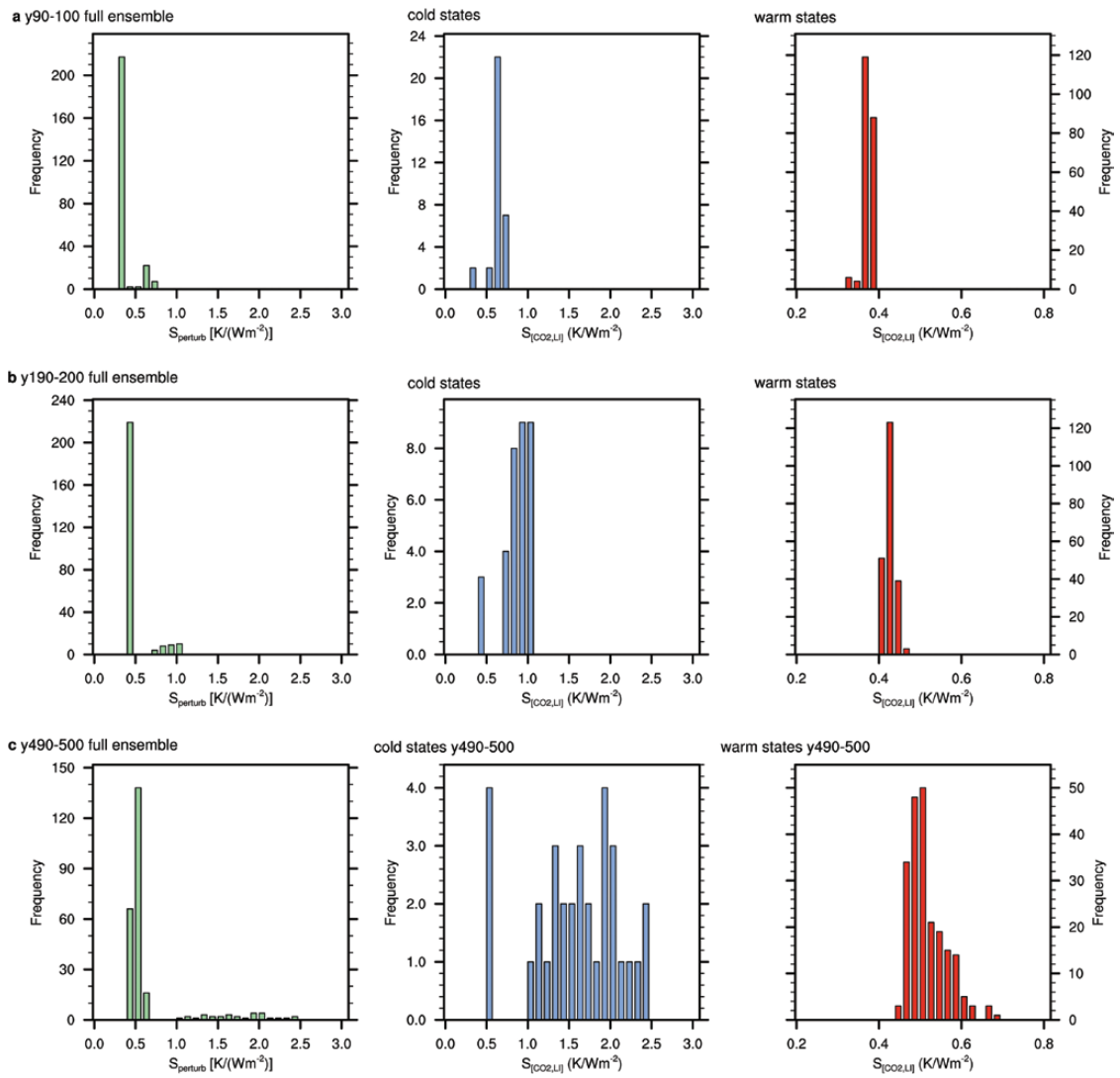

Figure 10. Climate sensitivity $S_{perturb}$ from perturbation experiments with dynamical CO_2. Shown are distributions of $S_{perturb}$ [cf. equation (10)] at different times after the perturbation. The ensemble starts from 250 initial conditions taken from the glacial–interglacial time series (500 kyr, shown in Figure 2a). The right panel in each plot shows the full ensemble, the middle panel shows only those initial states that have sea ice (classified as *C*) and the right panel shows the warm initial states without sea ice (classified as *W*). (**a**) $S_{perturb}$ after 100 years; (**b**) $S_{perturb}$ after 200 years; (**c**) $S_{perturb}$ after 500 years.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

4. Conclusions

In this paper, we have considered climate sensitivity as a local property of a climate attractor, in particular it is a property of a projection of a measure of this attractor on the (T,R) plane. This naturally leads to distributions of climate sensitivity for every radiative forcing, and if the attractor shows different regimes of special climate dynamics, state dependence of climate sensitivity can be explained in terms of regimes. We have explored this in a phenomenological Earth system model with the aim to test how climate sensitivity derived from palaeoclimate records might be compared to model-derived counterparts. Conceptually, climate sensitivity is defined differently in these two situations; while palaeoclimate time series reflect trajectories on the climate attractor, in model simulations generally perturbations away from the attractor are applied. Moreover, climate models include only a limited amount of processes (usually the slower processes are fixed, as is the carbon cycle), which means that a different attractor may be explored by the models.

Clearly, we cannot expect to get reliable quantitative conclusions about the distribution of ECS from the low-order conceptual model used for this study. Many important processes in the climate system (such as the impact of T on cloud formation) are absent from the model, which was constructed in Gildor and Tziperman (2001, 2002) with the aim of explaining ice age pacing rather than the link between T and CO_2. Even those processes that are included are open to debate; for example the sea ice cover changes in the model are 1.5 times larger than suggested by proxy data (Köhler *et al*., 2010), while Northern Hemisphere land ice cover changes are smaller (Fig. 2b). Moreover, the climate sensitivity derived from the model is higher during glacial periods because the fast sea ice-albedo feedback is stronger in those regimes. Proxy data suggest, however, higher climate sensitivity during warm periods (von der Heydt *et al*., 2014; Köhler *et al*., 2015), most likely because a combination of other fast feedbacks (such as water vapour, cloud feedbacks etc.) may be stronger during warm climates.

Nonetheless, even for this model, the presence of variability on a number of timescales and regimes within the attractor gives clear and non-trivial dependence of sensitivity on regime. This suggests that it could be useful to think of the unperturbed climate sensitivity (which can be determined from palaeoclimate data) as a property of the 'climate attractor'. For a perturbed system (we have considered instantaneously doubled CO_2), which is the normal approach in climate models, this is still useful once an initial transient has decayed. This transient will depend in particular on ocean heat uptake, though also on carbon cycle and biosphere processes that act on timescales roughly equivalent with the forcing timescale. In the case of a regime shift (either natural or induced by perturbation), the spread in climate sensitivity becomes very large. If the climate system has more than one attractor, the perturbed system may clearly evolve to a completely different set of states than the original attractor—a situation that does not occur in the climate model used here. In less extreme cases, we cannot rule out very long transients (associated with slow feedbacks) for some perturbations.

In most climate sensitivity studies, feedback processes are considered except those related to the carbon cycle. In the history of climate, those processes are active, however, on many different timescales. In our conceptual model, we have included the part of the carbon cycle that is related to the soft-tissue biological pump in the oceans and air-sea CO_2 exchange. The resulting CO_2 variations in the model's glacial–interglacial cycles are in the range of the observed glacial to interglacial CO_2 changes and amplify the glacial–interglacial cycle while they are not necessary to generate those cycles. Accordingly, when exploring climate sensitivity from perturbation experiments with the same model, we have instantaneously doubled CO_2 and kept the model's carbon cycle active. This procedure ensures that the perturbation experiments eventually return to the same attractor as the unperturbed system. Such perturbations (illustrated in Fig. 1b,c) are not normally applied in climate models used for climate predictions (IPCC, 2013), where climate sensitivity is derived from model simulations considering prescribed, non-dynamic atmospheric CO_2.

In our conceptual model, we have also examined climate sensitivities from a classical climate model perturbation (not shown); CO_2 is doubled within the first 30 years of the simulation and kept fixed afterwards for 200 years. In this case, we find significantly lower sensitivities and smaller spread than for $S_{perturb}$ obtained from doubling CO_2 with dynamic CO_2. This emphasizes the importance of including dynamic carbon cycle processes into climate projections. In this model, this supports the idea that the future observed climate response may indeed be larger than the (concentration-driven) model predicted one. However, the carbon cycle includes more timescales and processes than considered here in this simple model. For example, processes related to the ocean-seafloor system include carbonate compensation and silicate weathering, which act on much longer timescales and have been suggested to be responsible for a mean atmospheric lifetime of anthropogenic CO_2 of 30–35 kyr (Archer, 2005). Such processes may be responsible for a climate response larger than the model-determined, concentration-driven response, but at this moment, we cannot exclude other potentially negative feedback processes arising from the complete carbon cycle response.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

When deriving climate sensitivity from palaeoclimate records, it is important to take account of potential state dependence and different climate regimes before drawing conclusions on the ECS distribution that may be relevant for future climate evolution. For the conceptual model we consider, the long tail in the ECS distribution from the unperturbed (palaeoclimate) time series mostly results from the cross-comparison of states within different regimes (*CW*/*WC*). Similarly, the applied perturbation in the model always induced a regime transition and consequently large ECS values if the initial condition was in the *C* regime, but not in the *W* regime. In the context of our model, these high ECS values would not be relevant for the present climate continuing the current regime. On the other hand, if the present climate is in a regime that is susceptible to a regime shift (either natural or due to anthropogenic 'perturbation'), very large ECS values may be possible and indeed relevant. By studying data and models of warmer-than-present climates in the palaeorecord, we may be able to achieve information on potentially warmer climate regimes existing for perturbed versions of the climate attractor.

Acknowledgements

This work was carried out under the program of the Netherlands Earth System Science Centre (NESSC), financially supported by the Ministry of Education, Culture and Science (OCW) in the Netherlands. AH thanks CliMathNet (sponsored by EPSRC) for travel support to meetings that facilitated this work. We thank the Lorentz Center in Leiden for organizing a 'Workshop on Climate Variability: From Data and Models to Decisions' in 2014 where these ideas were first discussed, and the EU ITN 'CRITICS' for providing a further opportunity to discuss this research.

Appendix A. Model equations

A.1. Ocean and sea ice

The ocean consists of two layers of four meridionally oriented boxes, where the polar boxes extend from 45° to the pole and the equatorial boxes from the equator to 45°, with meridional lengths L_1, L_2, L_3, L_4 the same as the atmopsheric boxes. All tracers such as temperature T, salt S and biogeochemical variables are averaged over the two equatorial boxes, such that in fact the dynamics is determined by only three meridional boxes. The two vertical layers have thicknesses D_{upper} and D_{lower}, respectively. The ocean model dynamics includes a simple frictional horizontal momentum balance, is hydrostatic and mass conserving:

$$\begin{aligned} 0 &= -\frac{1}{\rho_0}\frac{\partial p}{\partial z} - \frac{g}{\rho_0}\rho \\ 0 &= -\frac{1}{\rho_0}\frac{\partial p}{\partial y} - rv \\ 0 &= \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \end{aligned} \tag{A.1}$$

Here, (y,z) are the meridional and vertical coordinates and (v,w) the corresponding flow velocities, respectively. p is the pressure, g the gravitational constant, ρ_0 a reference density and r a friction coefficient. In each box, temperature T and salinity S determine the density via the full non-linear equation of state as recommended by UNESCO (1981). Temperature and salinity are determined by the following balances:

$$\begin{aligned} \frac{\partial T}{\partial t} + \frac{\partial (vT)}{\partial y} + \frac{\partial (wT)}{\partial z} &= K_h \frac{\partial T}{\partial y} + K_v \frac{\partial T}{\partial z} + Q_T^{atm} + Q_T^{seaice} \\ \frac{\partial S}{\partial t} + \frac{\partial (vS)}{\partial y} + \frac{\partial (wS)}{\partial z} &= K_h \frac{\partial S}{\partial y} + K_v \frac{\partial S}{\partial z} + Q_S^{atm} + Q_S^{seaice} + Q_S^{landice} \end{aligned} \tag{A.2}$$

where K_h and K_v are horizontal and vertical diffusion coefficients, respectively. As in Gildor *et al.* (2002), the vertical mixing of any tracer tr (e.g. temperature, salinity or ocean CO_2) in the southern polar box is dependent on the vertical stratification:

$$K_v^0 (\sigma_{t_{deep}} - \sigma_{t_{surface}})^{-1} (tr_{deep} - tr_{surface}), \tag{A.3}$$

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

where $(\sigma_{t_{deep}} - \sigma_{t_{surface}}) \sim d\rho/dz$. In addition, upper and lower bounds of 280 and 1 Sv are imposed on the vertical mixing rates $K_v^0(\sigma_{t_{deep}} - \sigma_{t_{surface}})^{-1}$. Vertical mixing rates between the other surface and deep boxes are set constant, 0.25 Sv for the two equatorial boxes and 5 Sv for the northern polar box. The meridional overturning circulation is treated in the same way as in Gildor *et al.* (2002), with the upwelling through the southern polar box set to a fixed value of 16 Sv and the downwelling through the northern polar box determined by the meridional density gradient between the northern equatorial and polar ocean boxes.

The Q terms in the above equations are fluxes from other components of the climate model: Q_T^{atm} is the atmosphere–ocean heat flux due to sensible, latent and radiative fluxes:

$$Q_T^{atm} = \frac{\rho_0 C_{pw} D_{upper}}{\tau}(\theta - T)(f_{ow} + f_{si}\frac{\gamma}{D_{seaice}}), \tag{A.4}$$

where C_{pw} is the heat capacity of water, θ the temperature of the atmospheric box above and γ the insolation effect of a layer sea ice of thickness D_{seaice}. f_{ow} and f_{si} are the fractions of the ocean that are open water and sea ice covered, respectively, with $f_{ow} = 1 - f_{si}$. The timescale τ is chosen such that the ocean heat transport into the northern polar atmopsheric box is 2.3 PW during interglacial periods as in Gildor and Tziperman (2001). Precipitation P and evaporation E are converted into an equivalent salt flux:

$$Q_S^{atm} = -(P - E)S_0, \tag{A.5}$$

with S_0 a reference salinity. Heat and salt fluxes due to sea ice formation or melting are formulated as:

$$\begin{aligned} Q_T^{seaice} &= \frac{\rho_0 C_{pw} V_{ocean}}{\tau_{seaice}}(T^{seaice} - T), \\ Q_S^{seaice} &= \frac{Q_T^{seaice}}{\rho_{seaice} L_f} S_0, \end{aligned} \tag{A.6}$$

where V_{ocean} is the volume of the ocean box, $T^{sea-ice}$ is the temperature threshold where sea ice forms, L_f is the latent heat of fusion, ρ_{seaice} is the density of sea ice and τ_{seaice} is a short timescale to ensure that the ocean temperature remains close to the freezing temperature as long as sea ice is present. Sea ice is assumed to grow in area with an initial thickness of 3 and 1.5 m in the northern and southern polar boxes, respectively, until the whole box is covered. The volume of sea ice in the polar surface boxes V_{seaice} is given by:

$$\frac{dV_{seaice}}{dt} = \frac{Q_T^{seaice}}{\rho_{seaice} L_f} + P_{on-ice}. \tag{A.7}$$

P_{on-ice} is the amount of sea ice forming due to atmopsheric precipitation falling on the ocean area covered with sea ice.

A.2. Atmosphere

The atmospheric model follows that used in Gildor *et al.* (2002), with four atmospheric boxes above the ocean boxes. The lower surface of each atmospheric box can be either land or ocean and both can be partly covered with (land or sea) ice. The box-averaged potential temperature is calculated from the energy balance of the box, balancing incoming solar radiation (with a box albedo determined from the relative fraction of each lower surface type in the box), outgoing long-wave radiation at the top of the atmosphere, air–sea heat flux and meridional atmospheric heat transport. In each atmopsheric box, the temperature θ is determined by the difference between the heat flux at the top of the atmosphere F_{top} and at the surface $F_{surface}$, following the equation:

$$\begin{aligned} \frac{\partial \theta}{\partial t} &= \frac{2^{R/C_p} g}{P_0 C_p}\left[(F_{top} - F_{surface}) + (F_{merid}^{in} - F_{merid}^{out})\right] \\ &= \frac{2^{R/C_p} g}{P_0 C_p}\left[(H_{in} - H_{out} - Q_T^{atm}) + (F_{merid}^{in} - F_{merid}^{out})\right], \end{aligned} \tag{A.8}$$

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

where

$$\begin{aligned} H_{in} &= (1-\alpha_{surf})(1-\alpha_C)(1-q_{in}^{seaice})Q_{Solar} \\ H_{out} &= \left(\varepsilon - \kappa \ln\left(\frac{CO_2}{CO_{2,ref}}\right)\right)\sigma_B \theta^4, \end{aligned} \tag{A.9}$$

are the incoming and outgoing radiation terms at the top of the atmosphere, respectively. (R is the gas constant for dry air, C_p is the specific heat of the atmosphere at a constant pressure, P_0 a reference pressure, σ_B the Stefan-Boltzmann constant and g the gravitational acceleration.) The incoming solar radiation Q_{Solar} for each box is assumed to vary with season and due to orbital variations as in Gildor and Tziperman (2000). Furthermore, Q_{Solar} is reduced by a constant cloud albedo term α_C and a part q_{in}^{seaice} that is directly used to melt sea ice; where sea ice exists, 15% of the incoming short-wave radiation is used to melt sea ice and does not enter the radiation balance of the atmosphere (Gildor *et al.*, 2002). α_{surf} is the surface albedo of the box and is determined by the fraction of sea ice, land ice, land surface and ocean surface in that box:

$$\alpha_{surf} = f_L(1-f_{LI})\alpha_L + f_I f_{LI}\alpha_{LI} + f_O(1-f_{SI})\alpha_O + f_O f_{SI}\alpha_{SI} \tag{A.10}$$

Here, f_L, f_{LI}, f_O and f_{SI} correspond to the fraction of land, land ice, ocean and sea ice, respectively, and α_L, α_{LI}, α_O and α_{SI} to the corresponding albedos of each surface type. The outgoing radiation depends on a mean emissivity of the box ε and a term depending on the atmospheric CO_2 concentration. Here, κ is chosen (Gildor *et al.*, (2002)) such that a doubling of CO_2 will cause a radiative forcing of 4 Wm^{-2}. $F_{merid}^{in} - F_{merid}^{out}$ is the net heating due to meridional heat fluxes between the atmospheric boxes. Meridional heat transport between boxes is calculated as:

$$F_{merid} = K_\theta \nabla \theta, \tag{A.11}$$

where the coefficient K_θ is chosen such that the meridional heat transport between the two northern boxes is 2.2 PW during interglacial periods (Gildor and Tziperman, 2001). No net heat flux is assumed over land and land ice; therefore, $F_{surface}$ includes only the ocean–atmosphere heat exchange.

The meridional moisture transport F_{Mq} between the atmospheric boxes is parameterized as:

$$F_{Mq} = K_{Mq}|\nabla \theta| q, \tag{A.12}$$

where q is the humidity of the box. A constant relative humidity is assumed, with the saturation humidity at temperature θ calculated from an approximate Clausius–Clayperon equation:

$$q = 0.7 \cdot A \cdot e^{B/\theta}. \tag{A.13}$$

Over land ice in the polar boxes, another source of precipitation is the local evaporation of that part of the ocean box that is not covered by sea ice, with flux:

$$F_q = K_q f_{ow} q. \tag{A.14}$$

The total precipitation in each box is then given by

$$P - E = -\nabla \cdot (F_{Mq} + F_q). \tag{A.15}$$

Precipitation falling over land or sea ice is assumed to turn into additional ice.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

A.3. Land ice

The equations for the land ice sheets follow those of (Gildor and Tziperman, 2001), with the mass balance

$$\frac{dV_{ice-sheet}}{dt} = LI_{source} - LI_{sink}\,. \tag{A.16}$$

The source term L_{source} depends on the amount of precipitation falling over existing ice (or falling on the 0.3 poleward area of the box even if there is no glacier there):

$$LI_{source} = \frac{max\{0.3L_{area}, LI_{area}\}}{box_{area}}(P-E), \tag{A.17}$$

where L_{area} is the land area in the box, LI_{area} the ice sheet area and box_{area} the total area of the box.

The ice sheet can shrink as a consequence of ablation. The ablation term is assumed a constant C_{LI} (Gildor and Tziperman, 2001) plus a modulation by the summer Milankovitch forcing (Gildor and Tziperman, 2000):

$$LI_{sink} = C_{LI} + \gamma_{LI}(Solar_{June} - Solar_{ave,June}), \tag{A.18}$$

where $Solar_{June} - Solar_{ave,June}$ is the anomaly in summer insolation in this box relative to the average over the past 1 Myr. Southern Hemisphere ice sheets are assumed constant.

A.4. Biogeochemistry

In the ocean boxes, additional tracers are advected for total CO_2 (ΣCO_2), alkalinity (A_T) and phosphate PO_4. These are used to calculate atmopsheric pCO_2, see Gildor *et al.* (2002). The equations for the three biogeochemistry variables *Bio* in each ocean box follow:

$$\frac{\partial Bio}{\partial t} + \frac{\partial (vBio)}{\partial y} + \frac{\partial (wBio)}{\partial z} = K_h \frac{\partial Bio}{\partial y} + K_v \frac{\partial Bio}{\partial z} + S_{Bio}, \tag{A.19}$$

with additional source/sink terms S_{Bio} for these variables in the surface boxes:

$$\begin{aligned} S_{\Sigma CO_2} &= -R_C \times EP - RR \times EP + PV([CO_{2,a}] - [CO_{2,o}]) \\ S_{A_T} &= -2 \times RR \times EP + R_N \times EP \\ S_{PO_4} &= -EP, \end{aligned} \tag{A.20}$$

and in the deep boxes below:

$$\begin{aligned} S_{\Sigma CO_2} &= R_C \times EP + RR \times EP \\ S_{A_T} &= 2 \times RR \times EP - R_N \times EP \\ S_{PO_4} &= EP. \end{aligned} \tag{A.21}$$

EP and *RR* stand for export production and rain ratio, respectively, and R_C, R_N for the ratio $P:C$ and $P:N$ in particulate organic matter, respectively. $[CO_{2,a}]$ is the saturation concentration with regard to the partial pressure of CO_2 in the atmosphere, and $[CO_{2,o}]$ is the CO_2 concentration in the ocean. The flux of CO_2 between ocean and atmosphere $F_{CO_2} = PV([CO_{2,a}] - [CO_{2,o}])A_{openwater}$ is linearly related to the pCO_2 difference between the atmosphere and the surface ocean via a constant piston velocity *PV*, giving a timescale of 10 years for this gas exchange. For more details on the biogeochemistry module, see Gildor *et al.* (2002).

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

Appendix B. Climate sensitivity in the model

Climate sensitivity is determined from the energy balance of the Earth. For the conceptual model (Gildor and Tziperman, 2001), we can explicitly write the energy balance of the atmosphere and extract the different contributions to climate sensitivity. Averaged over all atmospheric boxes of the model the global mean temperature $T = \sum_{i=1}^{4} \frac{area_i}{area}\theta_i$ is determined by the difference between the heat flux at the top of the atmosphere F_{top} and at the surface $F_{surface}$ (see previous section), where $area_i$, $(i = 1,...,4)$, is the surface area of the four boxes and $area$ is the total surface area of the earth.

To access the contributions of the different forcings and feedbacks to the radiation balance, we split the global mean radiation terms into the different components due to solar radiation ($R_{[ins]}$), land ice ($R_{[LI]}$), sea ice ($R_{[SI]}$), outgoing long-wave radiation ($R_{[OLW]}$), CO_2 concentration ($R_{[CO_2]}$) and the radiation at the earth's surface ($R_{[surf]}$):

$$\frac{\partial T}{\partial t} = \frac{2^{R/C_p} g}{P_0 C_p} \left[R_{[ins]} + R_{[LI]} + R_{[SI]} + R_{[OLW]} + R_{[CO_2]} + R_{[surf]}\right] \tag{B.1}$$

The different contributions to the radiation balance can be expressed as:

$$R_{[ins]} = (1 - \alpha_C) Q_{solar} \tag{B.2}$$

$$R_{[LI]} = R_{[ins]} \sum_i \frac{area_i}{area} (f^i_L (1 - f^i_{LI})\alpha_L + f^i_L f^i_{LI} \alpha_{LI})(q^{seaice}_{in} - 1) \tag{B.3}$$

$$R_{[SI]} = -R_{[ins]} \sum_i \frac{area_i}{area} [q^{seaice}_{in} + (1 - q^{seaice}_{in})(f^i_O (1 - f^i_{SI})\alpha_O + f^i_O f^i_{SI} \alpha_{SI})] \tag{B.4}$$

$$R_{[OLW]} = -\sum_i \frac{area_i}{area} \varepsilon_i \sigma_B \theta_i^4 \tag{B.5}$$

$$R_{[CO_2]} = \sum_i \frac{area_i}{area} \kappa \ln \frac{pCO_2}{pCO_{2,ref}} \sigma_B \theta_i^4 \tag{B.6}$$

$$R_{[surf]} = -\sum_i \frac{area_i}{area} Q^i_{oa}. \tag{B.7}$$

When comparing two equilibrium climate states with global mean temperatures T_1 and T_2 (and $\Delta T = T_2 - T_1$), the radiation balance equation (42) reads:

$$0 = \Delta R_{[ins]} + \Delta R_{[LI]} + \Delta R_{[SI]} + \Delta R_{[OLW]} + \Delta R_{[CO_2]} + \Delta R_{[surf]}. \tag{B.8}$$

As we consider constant solar radiation and no changes in cloud albedo, $\Delta R_{[ins]} = 0$, and when we put all the forcing or slow feedbacks on the left-hand side and all fast feedback processes on the right-hand side, we obtain:

$$\Delta R_{[CO_2]} + \Delta R_{[LI]} = -\Delta R_{[OLW]} - \Delta R_{[SI]} - \Delta R_{[surf]}. \tag{B.9}$$

This finally leads to the expressions for the specific climate sensitivities

$$\begin{aligned}
S_{[CO_2]} &= \frac{\Delta T}{\Delta R_{[CO_2]}} = \frac{-\Delta T}{\Delta R_{[OLW]} + \Delta R_{[SI]} + \Delta R_{[surf]} + \Delta R_{[LI]}} \\
S_{[CO_2,LI]} &= \frac{\Delta T}{\Delta R_{[CO_2]} + \Delta R_{[LI]}} = \frac{-\Delta T}{\Delta R_{[OLW]} + \Delta R_{[SI]} + \Delta R_{[surf]}} \\
S_{[CO_2,LI,SI]} &= \frac{\Delta T}{\Delta R_{[CO_2]} + \Delta R_{[LI]} + \Delta R_{[SI]}} = \frac{-\Delta T}{\Delta R_{[OLW]} + \Delta R_{[surf]}}.
\end{aligned} \tag{B.10}$$

The last expression should approximate the sensitivity without feedbacks (i.e. only Planck feedback), $S_0 = (-4\varepsilon\sigma_B T^3)^{-1} \simeq 0.3$ K (W m^{-2})$^{-1}$. In the model, there is, however, one more radiation term due to the atmosphere–ocean heat exchange (ΔR_{surf}), which acts on fast to intermediate timescales. Therefore, $S_{[CO_2,LI,SI]}$ still slightly deviates from the Planck sensitivity.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017

References

Andrews T, Forster PM. CO_2 forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophys Res Lett* 2008; 35: L04802.

Archer DJ. Fate of fossil fuel CO_2 in geologic time. *J Geophys Res Oceans* 2005; 110: CO9S05.

Caballero R, Huber M. State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proc Natl Acad Sci USA* 2013; 110: 14162–67.

Charney JG. *Carbon Dioxide and Climate: A Scientific Assessment*. Washington, DC: National Academy of Science, 1979.

Chekroun MD, Simonnet E, Ghil M. Stochastic climate dynamics: random attractors and time-dependent invariant measures. *Physica D Nonlinear Phenom* 2011; 240: 1685–700.

Crucifix M. Does the Last Glacial Maximum constrain climate sensitivity? *Geophys Res Lett* 2006; 33: L18701.

Crucifix M. Oscillators and relaxation phenomena in Pleistocene climate theory. *it Philos Trans A Math Phys Eng Sci* 2012; 370: 1140–65.

Dijkstra HA, Viebahn JP. Sensitivity and resilience of the climate system: a conditional nonlinear optimization approach. *Commun Nonlinear Sci Numer Simul* 2015; 22: 13–22.

Ditlevsen PD, Ditlevsen OD. On the stochastic nature of the rapid climate shifts during the last ice age. *J Clim* 2009; 22: 446–57.

Ganopolski A, Rahmstorf S. Abrupt glacial climate change due to stochastic resonance. *Phys Rev Lett* 2002; 88: 038501-1–4.

Ghil M. A mathematical theory of climate sensitivity or, how to deal with both anthropogenic forcing and natural variability? In: Chang CP, Ghil M, Latif M, Wallace JM (eds). *Climate Change: Multidecadal and Beyond*. World Scientific Publ. Co./Imperial College Press, 2016, pp. 31–51.

Gildor H, Tziperman E. Sea ice as the glacial cycles' climate switch: role of seasonal and orbital forcing. *Paleoceanography* 2000; 15: 605–15.

Gildor H, Tziperman E. A sea ice climate switch mechanism for the 100-kyr glacial cycles. *J Geophys Res* 2001; 106: 9117–33.

Gildor H, Tziperman E, Toggweiler JR. Sea ice switch mechanism and glacial-interglacial CO_2 variations. *Glob Biogeochem Cyc* 2002; 16: 1032.

Gregory JM, Ingram WJ, Palmer MA *et al*. A new method for diagnosing radiative forcing and climate sensitivity. *Geophys Res Lett* 2004; 31: L03205.

Haywood AM, Dolan AM, Pickering SJ *et al*. On the identification of a Pliocene time slice for data–model comparison. *Philos Trans R Soc A Math Phys Eng Sci* 2013; 371: 20120515.

IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY: Cambridge University Press, 2013.

Knutti R, Hegerl GC. The equilibrium sensitivity of the earth's temperature to radiation changes. *Nat Geosci* 2008; 1: 735–43.

Köhler P, Bintanja R, Fischer H *et al*. What caused earth's temperature variations during the last 800,000 years? Data-based evidence on radiative forcing and constraints on climate sensitivity. *Quat Sci Rev* 2010; 29: 129–45.

Köhler P, de Boer B, von der Heydt AS, Stap LB, van de Wal RSW. On the state-dependency of the equilibrium climate sensitivity during the last 5 million years. *Climate Past* 2015; 11: 1801–23.

Lunt DJ, Haywood AM, Schmidt GA *et al*. Earth system sensitivity inferred from Pliocene modelling and data. *Nat Geosci* 2010; 3: 60–4.

Martinez-Boti MA, Foster GL, Chalk TB *et al*. Plio-Pleistocene climate sensitivity evaluated using high-resolution CO_2 records. *Nature* 2015; 518: 49–54.

Rohling EJ, Sluijs A, Dijkstra HA *et al*.; PALAEOSENS Project Members. Making sense of palaeoclimate sensitivity. *Nature* 2012; 491: 683–91.

Scheffer M, Brovkin V, Cox PM. Positive feedback between global warming and atmospheric CO_2 concentration inferred from past climate change. *Geophys Res Lett* 2006; 33: L10702.

Schulz M. The tempo of climate change during Dansgaard-Oeschger interstadials and its potential to affect the manifestation of the 1470-year climate cycle. *Geophys Res Lett* 2002; 29: 1002.

Schwartz SE. Determination of earth's transient and equilibrium climate sensitivities from observations over the twentieth century: strong dependence on assumed forcing. *Surv Geophys* 2012; 33: 745–77.

Senior CA, Mitchell JFB. The time dependence of climate sensitivity. *Geophys Res Lett* 2000; 27: 2685–88.

UNESCO. *10th Report of the Joint Panel on Oceanographic Tables and Standards*. Technical Report 36, UNESCO Technical Paper in Marine Science, 1981.

von der Heydt AS, Köhler P, van de Wal RSW, Dijkstra HA. On the state dependency of fast feedback processes in (paleo) climate sensitivity. *Geophys Res Lett* 2014; 41: 6484–92.

Yoshimori M, Hargreaves JC, Annan JD, Yokohata T, Abe-Ouchi A. Dependency of feedbacks on forcing and climate state in physics parameter ensembles. *J Climate* 2011; 24: 6440–55.

Downloaded from https://academic.oup.com/climatesystem/article-abstract/1/1/dzx001/3068110
by University of Exeter user
on 30 November 2017